



별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto
is a true copy from the records of the Korean Intellectual
Property Office.

출 원 번 호 : 10-2003-0063517
Application Number

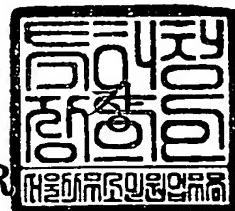
출 원 년 월 일 : 2003년 09월 15일
Date of Application SEP 15, 2003

출 원 인 : 한국전자통신연구원
Applicant(s) Electronics and Telecommunications Research Insti



2003 년 10 월 28 일

특 허 청
COMMISSIONER



【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【제출일자】	2003.09.15
【발명의 명칭】	규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 장치 및 방법과 그 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체
【발명의 영문명칭】	Hybrid automatic translation Apparatus and Method by combining Rule-based method and Translation pattern method, and The medium recording the program
【출원인】	
【명칭】	한국전자통신연구원
【출원인코드】	3-1998-007763-8
【대리인】	
【성명】	권태복
【대리인코드】	9-2001-000347-1
【포괄위임등록번호】	2001-057650-1
【대리인】	
【성명】	이화익
【대리인코드】	9-1998-000417-9
【포괄위임등록번호】	1999-021997-1
【발명자】	
【성명의 국문표기】	노윤형
【성명의 영문표기】	ROH, Yoon-Hyung
【주민등록번호】	730303-1655110
【우편번호】	305-345
【주소】	대전광역시 유성구 신성동 208-9 301호
【국적】	KR
【발명자】	
【성명의 국문표기】	최승권
【성명의 영문표기】	CHOI, Sung Kwon
【주민등록번호】	621219-1001217

【우편번호】 305-333
 【주소】 대전광역시 유성구 어은동 한빛아파트 113-1101
 【국적】 KR
 【발명자】
 【성명의 국문표기】 이기영
 【성명의 영문표기】 LEE,Kiyoung
 【주민등록번호】 691218-1117411
 【우편번호】 302-740
 【주소】 대전광역시 서구 만년동 초원아파트 107-1406
 【국적】 KR
 【발명자】
 【성명의 국문표기】 흥운표
 【성명의 영문표기】 HONG,Munpyo
 【주민등록번호】 691009-1041520
 【우편번호】 305-340
 【주소】 대전광역시 유성구 도룡동 383-2 과기원교수아파트 2동 102호
 【국적】 KR
 【발명자】
 【성명의 국문표기】 류철
 【성명의 영문표기】 RYU,Cheol
 【주민등록번호】 730828-1462918
 【우편번호】 302-723
 【주소】 대전광역시 서구 관저 2동 구봉마을아파트 512-1001
 【국적】 KR
 【발명자】
 【성명의 국문표기】 박상규
 【성명의 영문표기】 PARK,Sang Kyu
 【주민등록번호】 590413-1683310
 【우편번호】 305-503
 【주소】 대전광역시 유성구 송강동 한마을아파트 105-1305
 【국적】 KR

【발명자】

【성명의 국문표기】 김영길
 【성명의 영문표기】 KIM, Young Kil
 【주민등록번호】 680919-1120416
 【우편번호】 301-758
 【주소】 대전광역시 중구 오류동 삼성아파트 1-1005
 【국적】 KR

【발명자】

【성명의 국문표기】 김창현
 【성명의 영문표기】 KIM, Chang Hyun
 【주민등록번호】 681002-1231613
 【우편번호】 140-190
 【주소】 서울특별시 용산구 후암동 448-156(140-190)
 【국적】 KR

【발명자】

【성명의 국문표기】 서영애
 【성명의 영문표기】 SEO, Young-Ae
 【주민등록번호】 740209-2690711
 【우편번호】 305-330
 【주소】 대전광역시 유성구 지족동 열매마을 대우아파트 303동 1003호
 【국적】 KR

【발명자】

【성명의 국문표기】 양성일
 【성명의 영문표기】 YANG, SEONG IL
 【주민등록번호】 710407-1841111
 【우편번호】 305-345
 【주소】 대전광역시 유성구 신성동 두레아파트 110/1008
 【국적】 KR

【심사청구】

【취지】 특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사 를 청구합니다. 대리인
 권태복 (인) 대리인
 이화익 (인)

1020030063517

출력 일자: 2003/11/3

【수수료】

【기본출원료】	20	면	29,000	원
【가산출원료】	3	면	3,000	원
【우선권주장료】	0	건	0	원
【심사청구료】	11	항	461,000	원
【합계】			493,000	원
【감면사유】			정부출연연구기관	
【감면후 수수료】			246,500	원

【기술이전】

【기술양도】	희망
【실시권 허여】	희망
【기술지도】	희망
【첨부서류】	1. 요약서·명세서(도면)_1통

【요약서】**【요약】**

본 발명은 규칙 기반 방식에서의 모호성 문제와 번역 패턴 방식에서의 패턴 생성 및 커버리지 문제를 해결하기 위해 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 장치 및 방법과 그 기록 매체에 관한 것이다.

본 발명에 따른 자동번역 장치는, 입력 원문에 대해 형태소를 분석하고 품사를 결정하는 형태소 분석 및 태깅 수단; 태깅 결과에 대해 구문 분석을 수행하여 파싱 트리를 출력하는 구문분석 수단; 파싱트리에서 하위범주의 구 청킹 결과만을 추출하여 구문패턴을 생성하는 구문 패턴 생성 수단; 번역 패턴을 이용하여 상기 구문패턴에 대한 번역을 시도하는 구문패턴 번역 수단; 상기 구문패턴에 대한 번역패턴 매칭에 실패한 경우, 절구조 분석을 하는 절구조 분석 수단; 및 절구조 분석 결과에 따라 부분 구문패턴의 패턴번역을 수행하여 최종 번역결과를 출력하는 부분패턴 번역 수단;을 포함하는 것을 특징으로 한다

【대표도】

도 1

【색인어】

파싱, 청킹, 커버리지, 구문패턴, 부분 패턴번역, 절구조 분석

【명세서】

【발명의 명칭】

규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 장치 및 방법과 그 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체{Hybrid automatic translation Apparatus and Method by combining Rule-based method and Translation pattern method, and The medium recording the program}

【도면의 간단한 설명】

도 1은 본 발명에 따른 하이브리드 자동 번역 장치의 구성요소 및 처리 흐름을 보여주는 블록도.

도 2는 본 발명에 따른 구문 분석부의 구성 및 처리 흐름을 보여주는 블록도.

도 3은 본 발명에 따른 부분 패턴 번역 과정에 대한 처리 흐름도.

도 4는 본 발명에 따른 부분 패턴 번역 과정의 일 실시예.

<도면의 주요부분에 대한 부호의 설명>

101: 형태소 분석부 102: 태깅부

103: 구문 분석부 104: 구문 패턴 생성부

105: 구문 패턴 번역부 105-1: 부분 패턴 번역부

106: 절 구조 분석부 107: 번역 패턴 DB

【발명의 상세한 설명】

【발명의 목적】

【발명이 속하는 기술분야 및 그 분야의 종래기술】

- <10> 본 발명은 자동 번역 장치 및 방법과 그 기록매체에 관한 것이며, 보다 상세히는 종래의 규칙 기반 방식에서의 모호성 문제와 번역 패턴 방식에서의 패턴 생성 및 커버리지 문제를 해결하기 위해 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 장치 및 방법과 이를 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록 매체에 관한 것이다.
- <11> 종래의 규칙 기반 기계 번역 방법에서는, 특히 문장이 길어짐에 따라, 구문 분석이 가지는 애매성 폭발 및 대역 구문의 무제한 생성에 의해 속도 및 번역 성능이 저하되는 문제가 있었다.
- <12> 이를 해결하기 위한 것으로 번역패턴 기반의 자동번역 방법이 있으며, 이는 원문에서 미리 정해진 번역패턴을 발견하는 방법으로서 대역 구문의 무제한 생성을 방지하고 번역의 품질을 크게 향상시키는 장점이 있다.
- <13> 하지만, 종래의 번역패턴 기반의 자동번역 방법은, 태깅, 부분 파싱 등만으로는 번역을 위한 구문 패턴을 생성하는 데까지 발생하는 모호성을 처리하지 못하고 올바른 구문 패턴 자체를 생성하지 못함으로 인해 번역 패턴 기반의 장점을 발휘하는데 제한이 있었다.
- <14> 더욱이, 문장의 길이가 길어짐에 따라 구축해야 할 번역패턴의 수가 급격히 증가하게 되고, 번역패턴에 대한 매칭 성공률이 떨어져 심각한 커버리지 문제를 갖게 된다.
- <15> 또한, 이러한 커버리지 문제를 해결하기 위한 기존의 대표적인 장문 처리방법은 구문 분석을 하기 전에 장문을 분할하여 더 작은 단위로 나누어 처리하는 것인데, 기존의 장문 분할

방법은 구문 분석이 이루어지기 전의 제한된 정보를 가지고 수행함으로 인해 성능의 한계 및 부작용이 많았다.

【발명이 이루고자 하는 기술적 과제】

<16> 따라서, 본 발명은 상술한 종래의 문제점을 해결하기 위한 것으로서, 본 발명의 목적은 번역 패턴 방식에서 입력문장에 대한 구문 패턴을 구문 분석 결과에서 구 청킹(chunking) 결과만을 추출하여 생성함으로써, 규칙기반 방식의 모호성 문제를 피하면서 구문 패턴 생성의 정확성을 높이고, 또한 패턴 번역에 실패하는 경우, 절구조 분석만을 다시 수행하여 그 결과에 따라 부분패턴 번역을 수행함으로 번역 패턴 기반의 자동 번역에 있어 문장 길이가 길어짐에 따라 발생하는 번역의 커버리지 문제를 해결하여 높은 커버리지의 고품질 자동번역 결과를 생성할 수 있는 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 장치 및 방법과 그 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공하는데 있다.

<17> 상기 본 발명의 목적을 달성하기 위한 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 장치는, 입력 원문에 대해 형태소 분석을 수행하는 형태소 분석부; 상기 형태소 분석결과에 대해 각각의 품사를 결정하는 태깅부; 상기 태깅결과에 대해 구문 분석을 하고 파싱트리를 출력하는 구문 분석부; 상기 파싱트리에서 동사의 하위범주에 속하는 구들의 청킹 결과만을 추출하여 구문 패턴을 생성하는 구문패턴 생성부; 번역 패턴을 이용하여 상기 구문 패턴에 대한 번역을 수행하는 구문패턴 번역부; 상기 구문 패턴에 대한 번역 패턴 매칭에 실패한 경우, 그 구문에 대한 절 단위의 구조를 파악하는 절구조 분석부; 및 상기 절 구조 분석 결과를 참조하여 각 하위절에 대한 부분 구문패턴을 인식하고, 부분 번역 패턴을 이용하여 번역을 수행하는 부분패턴 번역부;로 구성된다.

<18> 또한, 상기 본 발명의 목적을 달성하기 위한 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 방법은, 입력 원문에 대해 형태소 분석 및 전처리 청킹을 수행하고, 그 분석결과에 대해 태깅하는 형태소 분석 및 태깅 단계; 상기 태깅결과에 대해 구문 분석을 하고 파싱트리를 출력하는 구문 분석 단계; 상기 파싱트리에서 동사의 하위법주에 속하는 구들의 청킹 결과만을 추출하여 구문 패턴을 생성하는 구문패턴 생성 단계; 번역 패턴을 이용하여 상기 구문 패턴에 대한 번역을 수행하는 구문패턴 번역 단계; 상기 구문 패턴에 대한 번역 패턴 매칭에 실패한 경우, 구문의 절 단위 구조를 분석하는 절구조 분석 단계; 및 상기 절 구조 분석 결과를 참조하여 번역 실패 노드의 하위절에 대한 부분 구문패턴을 생성하고, 그 부분 구문패턴에 대한 패턴 번역을 수행하고 이를 조합하여 최종 번역결과를 출력하는 부분패턴 번역 단계;로 이루어진다.

<19> 또한, 상기 부분패턴 번역 단계는, 상기 절 구조 분석 결과를 참조하여, 번역실패 노드의 하위절에 대한 부분 구문패턴을 생성하고, 그 부분 구문패턴에 대해 패턴 번역을 수행하며; 상기 부분 구문패턴의 번역결과를 문장 심볼 S로 치환하고, 그 패턴치환으로 축소된 구문패턴에 대해 패턴 번역을 수행하며; 상기 구문패턴 축소에 의한 패턴 번역이 실패할 경우, 각 구문 요소별로 번역을 수행하여 최종 번역결과를 생성하는; 것이 바람직하다.

<20> 또한, 상기 본 발명의 목적을 달성하기 위하여, 본 발명에 따른 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 방법을 실현할 수 있는 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공한다.

【발명의 구성 및 작용】

- <21> 이하, 본 발명에 따른 실시예를 첨부한 도면을 참조하여 상세히 설명하기로 한다.
- <22> 도 1은 본 발명에 따른 하이브리드 자동 번역 장치의 각 구성요소 및 처리 흐름을 보여주는 전체적인 블록구성이다.
- <23> 도 1에서, 본 발명의 자동번역 장치에 대한 전체적인 흐름을 살펴보면, 입력문장에 대해 형태소 분석 및 태깅(tagging)을 수행하고,(101,102) 태깅 결과로 들어온 입력 문장에 대하여 구문 분석(parsing)을 수행한다.(103) 그리고, 구문 분석 결과 생성된 파싱 트리로부터 구문 패턴을 생성한 후,(104) 번역 패턴에 의해 번역을 수행하게 된다.(105)
- <24> 여기에서, 구문 패턴은, 입력 문장에서 문장의 중심이 되는 동사(V), 조동사(X), 접속사(C) 등의 품사와 거기에 의존하는 구문요소로 이루어지는 전체 문장을 나타내는 패턴을 말한다. 또한, 구문요소로는 명사구(NP), 전치사구(PP), 형용사구(AP), 고립 전치사구(IPREP) 등이 있고, 각각은 n(명사구), p(전치사구), a(형용사구), i(고립전치사구)의 심볼로 표시한다.
- <25> 본 발명에서의 구문 패턴은 상기 품사나 구문요소로 이루어진 문장 단위의 패턴을 의미하는 것으로, 구 단위의 패턴을 사용하는 일반적인 패턴기반 방식의 번역에서의 패턴과 구별되는 것이다. 또한, 이러한 구문 패턴에 대응하는 대역문의 대역 구문패턴을 기술함으로써 입력 문장에 꼭 적절한 대역문 생성이 가능하도록 할 수 있는데, 이러한 문장 범위의 번역 정보를 담고 있는 구문 단위 패턴을 번역 패턴이라 한다. 이러한 번역 패턴에 의한 번역 방식은 철저한 구문구조를 파악해야만 번역성능이 보장되기 때문에 번역이 어려운 영어-한국어와 같은 이종언어간에 높은 성능을 발휘하게 된다.

- <26> 또한, 본 발명은, 상기 번역 패턴에 의한 번역에서 번역 패턴 매칭에 실패할 경우, 절구조 분석을 수행하고, (106) 절구조 분석 결과에 따라 부분패턴 번역을 수행하게 된다. (105-1)
- <27> 이와 같은 부분패턴 번역은, 문장전체에 대한 번역 패턴이 존재하지 않는 경우, 하위절 (sub-clause)에 해당하는 부분 구문 패턴으로 나누어 처리하고 그 결과를 연결하여 최종결과를 생성함으로 번역 패턴의 커버리지를 높이기 위함이다.
- <28> 이하에서는, 도 1 내지 도 4를 참조하여, 본 발명에 따른 자동번역 장치를 각 세부 블록 별로 보다 상세히 기술한다.
- <29> 도 1에서, 형태소 분석부(101)는, 입력되는 원문에 대해 형태소 분석 및 전처리 청킹 (chunking)을 수행한다. 전처리 청킹은 고유명사, 시간 부사구, 어휘 고정표현 등을 미리 묶어 줌으로써 문장의 길이를 줄이고 태깅 성능을 올릴 수 있게 한다.
- <30> 또한, 태깅부(102)는, 상기 형태소 분석에 대해 태깅을 수행하며, 그 태깅 결과는 태깅 자체 성능 및 파싱 효율성을 고려하여 각 단어에 대해 최적의 후보 2개를 출력한다. 따라서, 태깅만으로는 분별이 어려운 모호성이 있는 경우, 파싱을 통해 넓은 범위의 구문 분석 정보를 반영함으로 인한 태깅 성능의 향상을 기대할 수 있다.
- <31> 한편, 도 2는, 구문 분석부(103)의 세부 블록구성을 보여주는 도면이다.
- <32> 도 2에서, 구문 분석부(103)는 입력되는 두 개의 태깅 최적후보에 대해 파싱(parsing)을 수행하는(S201) 구성블록으로, 입력 문장의 길이가 특정값(N) 이상인 장문일 경우 문장 분할에 의한 파싱을 수행한다. 이때, 장문의 판정은 전처리 청킹이 된 상태에서의 문장 길이로 이루어 진다.
- <33> 본 발명에 따른 문장 분할에 의한 파싱은 다음과 같은 과정으로 이루어진다.

<34> 먼저, 문장의 구두점, 접속사, 관계사, 의문사 등의 분할점 구문단서를 바탕으로 다수의 문장 분할점 후보를 선정한 후, 선정된 후보들 중에서 각 분할문 양쪽에 본동사(즉, 시제를 가지고 있는 동사)가 존재하는지 여부 및 분할문의 길이를 고려하여 2 내지 3개의 분할점 후보를 선정한다.(S202)

<35> 그리고, 각 후보별로, 그 분할점에 의한 분할문들에 대해 파싱을 수행한다.(S203) 만일, 분할문 자체가 장문일 경우는 상기 S202 단계 및 S203 단계를 재귀적으로 적용하여 파싱을 수행한다. 이와 같이 분할문 자체의 길이가 특정값 이상인 분할문에 대해 다시 장문 분할을 재귀적으로 수행함으로써 임의의 장문에 대해서도 원하는 만큼 분할을 수행할 수 있다.

<36> 그리고, 각 분할문 파싱결과에 파싱 가중치를 적용하여 가중치가 높은 최적의 분할점을 선정하고, 선정된 분할점에 따른 파싱결과 및 파싱트리를 출력한다.(S204)

<37> 또한, 삽입절과 같이 분할되서는 안되는 지점을 찾기 위해서는 매우 넓은 범위의 문맥과 깊은 분석을 필요로 하는데, 본 발명은 각 후보별로 파싱을 수행한 후 최종 분할점을 결정하기 때문에 최적의 분할점을 보다 정확하게 판정할 수 있다.

<38> 다음은 아래의 입력문장(영문)에 대한 본 발명에 의한 문장 분할에 의한 파싱의 일 실시 예를 나타낸다.

<39> [입력문장]: "We're told to look for an announcement under which the Russians would temporarily participate in the NATO command structure while the political leaders, including the two presidents when they speak today, try to work out the arrangements for a much broader Russian participation in the peacekeeping force."

<40> [분할점 후보]: ... in the NATO command structure /while the political leaders, including the two presidents /when they speak today, try to

<41> [각 분할점 후보별 분할문]

<42> while: (We're told to look for ... NATO command structure) (while the political leaders, including the two presidents when they speak today, try to ... the peacekeeping force.)

<43> when: (We're told to look for ... NATO command structure while the political leaders, including the two presidents) (when they speak today, try to ... in the peacekeeping force.)

<44> 분할점 후보 'when'의 경우, 그 분할문 "We're told to look for an announcement under which the Russians would temporarily participate in the NATO command structure while the political leaders, including the two presidents" 은 비문(abnormal sentence)이므로, 파싱 가중치에 의해 'when'은 분할점 후보에서 제외된다.

<45> [최종 선정된 분할문의 파싱 결과]

<46> (S (NP We) (VP 're (VP told (TOINF (VP to (VP look_for (NP an announcement) (PP under)))))) (SBAR (WHNP which) (SS (NP the Russians) (VP would temporarily (VP participate (PP in (NP the NATO command structure)))))))

<47> (S (NP (NP the political leaders) -COMMA- (PP including (NP (NP the two presidents) (SBAR (WHADVP when) (SS (NP they) (VP speak today)))))) -COMMA-) (VP try (TOINF to (VP work_out) (NP the arrangements) (PP for (NP (NP a (ADJP much broader) Russian participation) (PP in (NP the peacekeeping force)))))))

<48> 구문 패턴 생성부(104)는 상기 최종 선정된 분할점 후보에 대한 파싱 트리에서 NP, AP, PP, IPREP와 같이 동사의 하위범주에 속하는 구(phase)들의 청킹 범위를 인식함으로써 구문패턴을 추출한다.

<49> 본 발명에서 동사의 하위 범주라함은 구문 트리상의 NP, AP, PP, IPREP 중에서 동사에 의존하는 구를 말한다. 구문 트리 상에서 주로 상위로 갈수록 모호성이 증가하기 때문에, 본 발명은 이와 같이 하위범주의 구 청킹 결과만으로 구문패턴을 추출함으로써 구문분석의 모호성 문제를 줄일 수 있다.

<50> 다음은 상기 입력 예문에 대한 구 청킹 추출 결과 및 구문 패턴이다.

<51> [구 청킹 추출 결과]

<52> (NP We) 're told (IPREP to) look_for (NP an announcement) (IPREP under) which (NP the Russians) would temporarily participate (PP in the NATO command structure)

<53> (NP the political leaders) -COMMA- (PP including the two presidents) when (NP they) speak today -COMMA- try (IPREP to) work_out (NP the arrangements) (PP for a much broader Russian participation in the peacekeeping force)

<54> [구문 패턴]: nViVniCnVpCnTpCnVTViVnp

<55> 위에서 보면, 'while'은 실제로 'under which'의 관계절 안의 접속사로서, 분할되어서는 안 되는 분할점이다. 따라서, 'while'에 의해 분할된 상태로 종래 방식에 따라 번역을 수행하게 된다면 잘못된 번역 결과를 생성하게 될 것이다. 즉, 종래 방식의 경우, 분할점의 선정에 의해 번역결과가 결정되어버리게 된다.

<56> 하지만, 본 발명은, 선정된 파싱 결과 중에서 하위법주의 구단위 청킹 결과만을 사용하여 구문패턴을 추출하기 때문에, 분할점의 선정이 구문패턴 결과에 큰 영향을 미치지 않게 되고, 올바른 절구조는 다시 절구조 분석을 통해 얻어지게 된다. 결과적으로, 문장 분할의 실패에 의한 위험성이 감소하게 된다.

<57> 한편, 구문 패턴 번역부(105)는 상기 추출된 구문 패턴에 대해 번역 패턴 DB(107)에서 패턴 매칭을 수행한다. 만일, 전 구문에 대한 번역패턴 매칭이 성공하면, 그 번역 패턴에 의해 번역을 수행하고 결과를 출력한다.

<58> 하지만, 상기 구문 패턴에 대한 번역패턴의 매칭이 실패할 경우, 절구조 분석부(106)는 그 구문 패턴에 대해 절구조 분석을 수행한다.

<59> 절구조 분석은 문장내 본동사를 포함하는 절 단위의 구조를 파악하는 것으로, 입력 예문에 대해 다음과 같은 절구조 분석결과가 나오게 된다.

<60> [절구조 분석 결과]

<61> (s nViVniC(s (s nVp)C(s nT(p pC(s nV))TViVnp)))

<62> 그리고, 부분 패턴 번역부(105-1)에서, 절구조 분석 결과를 토대로 부분 번역 패턴을 이용한 번역을 수행한다.

<63> 도 3은 본 발명에 따른 패턴 번역의 처리 흐름을 보여준다.

<64> 도 3에서, 본 발명의 구문 패턴 번역은, 먼저 입력되는 구문 패턴에 대해 번역 패턴 매칭 및 번역을 수행한다.(S301) 이때, 패턴 번역에 성공하면 그 번역결과를 출력하고 종료한다.

<65> 하지만, 구문 패턴 번역에 실패할 경우, 절구조 분석을 수행하고, 그 절구조 분석 트리로부터 현재의 하위노드에 해당하는 범위에 대한 부분 구문 패턴을 생성한다. 이때, 관계절이나 의문사절 등의 경우에는 이동된 원래의 구문요소를 복원시켜 기존의 번역 패턴에 의해 번역이 될 수 있도록 문장 복원을 수행한다.

<66> 그리고, 상기 생성된 하위의 부분 구문 패턴에 대해 상기 패턴 번역 DB(107)를 참조하여 패턴 번역을 수행한다.(S302) 이때, 그 부분 구문 패턴에 대한 패턴 번역에 실패할 경우, 다시 절구조 분석 결과를 참조하여 그 하위 절에 대한 부분 패턴 번역을 수행하게 된다.

<67> 그리고, 각 하위절에 해당하는 부분 구문 패턴에 대한 번역 결과가 나오면, 해당 범위의 번역결과를 담고 있는 문장 심볼 S로 치환하고, 그 패턴치환으로 reduce된 구문 패턴에 대해 번역 패턴 매칭 및 번역을 수행함으로써 최종 번역결과를 생성하게 된다.(S303)

<68> 만약, 상기의 reduce된 구문 패턴에 의한 번역도 실패할 경우, NP, Verb, S(번역된 하위절), AP 등과 같은 구문패턴을 이루는 각 구문요소별로 번역을 수행하고, 이들을 조합하여 최종 번역결과를 생성한다.(S304)

<69> 한편, 도 4는 상기 입력예문에 대한 절 구조 분석 결과 및 부분패턴 번역의 일 실시예를 나타낸다.

<70> 도 4에서, 먼저 s1에 대한 패턴 번역을 시도하고 이에 실패할 경우, 그 절구조 분석결과로부터 하위절인 s2를 인식하고 1.1)에서 s2의 번역을 시도한다. 이때, s2에 대한 번역에 성공하게 되면 1.2)에서와 같이 reduce된 구문 패턴에 대해 번역을 함으로써 전체 번역이 이루어지게 된다.

<71> 만약, s2의 부분 구문 패턴에 대한 직접 번역이 실패할 경우, 다시 절구조 분석 결과에서 그 하위절인 s3, s4를 인식한 후, 1.1.1), 1.1.2), 1.1.3)에서와 같이 하위 부분패턴 번역을 시도하며, 하위 번역 패턴에 대해서도 패턴 번역이 실패하는 경우 그 하위에 대해 동일한 과정을 반복하게 된다. 또한, 최종 하위절에 대한 패턴 번역에 실패할 경우는, 각 구문요소별로 번역을 시도한다.

<72> 본 발명은 이와 같이 하향식으로 부분패턴 번역을 수행하기 때문에, 만일 절구조 분석상에 오류가 있다 하더라도 그 상위의 구조에서 번역 패턴이 존재하게 되면 번역 패턴에 의해 올바른 번역이 수행되므로, 절구조 분석상의 오류에 의한 부작용을 최소화할 수 있다.

<73> 또한, 전체 구문에 대한 번역 패턴이 없을 경우 하위절의 부분구문 패턴 및 reduced 구문패턴으로 매칭하기 때문에, 매칭될 패턴의 길이가 줄어들게 되고 번역 패턴의 커버리지를 효과적으로 높일 수 있다.

【발명의 효과】

<74> 상술한 바와 같이 본 발명에 의하면, 구조 분석의 처리 단위를 구단위와 절단위로 구분하고, 구문 분석 결과에서 구단위 결과만 추출함으로써, 구문분석의 모호성 문제, 문장분할의 부작용 문제를 최소화하고, 번역 패턴 매칭을 위한 구문 패턴의 정확성을 높일 수 있다.

<75> 또한 절구조 분석 결과로부터 하향식 방식으로 부분패턴 번역을 수행함으로써, 높은 커버리지의 고품질 번역결과를 얻을 수 있다.

<76> 이상에서 설명한 것은 본 발명에 따른 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 장치 및 방법을 실시하기 위한 하나의 실시예에 불과한 것으로서, 본 발명은 상기한 실시예에 한정되지 않고, 이하의 특허청구의 범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 분야에서 통상의 지식을 가진 자라면 누구든지 다양한 변경 실시가 가능한 범위까지 본 발명의 기술적 정신이 있다고 할 것이다.

【특허청구범위】**【청구항 1】**

입력 원문에 대해 형태소 분석을 수행하는 형태소 분석부;

상기 형태소 분석결과에 대해 각각의 품사를 결정하는 태깅부;

상기 태깅결과에 대해 구문 분석을 하고 파싱트리를 출력하는 구문 분석부;

상기 파싱트리에서 동사의 하위범주에 속하는 구들의 청킹 결과만을 추출하여 구문 패턴을 생성하는 구문패턴 생성부;

번역 패턴을 이용하여 상기 구문 패턴에 대한 번역을 수행하는 구문패턴 번역부;

상기 구문 패턴에 대한 번역 패턴 매칭에 실패한 경우, 그 구문에 대한 절 단위의 구조를 파악하는 절구조 분석부; 및

상기 절 구조 분석 결과를 참조하여 각 하위절에 대한 부분 구문패턴을 인식하고, 부분 번역 패턴을 이용하여 번역을 수행하는 부분패턴 번역부;로 구성되는 것을 특징으로 하는 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 장치.

【청구항 2】

제 1항에 있어서, 상기 형태소 분석부는, 입력 원문에 대한 형태소 분석 시 전처리 청킹을 수행하는 것을 특징으로 하는 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 장치.

【청구항 3】

제 1항에 있어서, 상기 태깅부는, 그 태깅 결과로서 최적의 후보 2개를 출력하여 상기 구문 분석부로 제공하는 것을 특징으로 하는 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 장치.

【청구항 4】

제 1항에 있어서, 상기 구문 분석부는, 입력 문장의 길이가 특정값 이상인 장문일 경우, 분할점 구문단서, 본동사 존재 여부, 분할문의 길이를 바탕으로 2 내지 3개의 분할점 후보를 선정하고;

각 후보별로 그 분할문들에 대한 파싱을 수행하며;
각 분할문 파싱결과에 파싱 가중치를 적용하여 최적의 분할점을 선정하고, 선정된 분할점에 따른 구문 파싱결과를 출력하는; 것을 특징으로 하는 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 장치.

【청구항 5】

제 1항에 있어서, 상기 부분패턴 번역부는,
상기 절 구조 분석 결과를 참조하여, 번역실패 노드의 하위절에 대한 부분 구문패턴을 생성하고, 그 부분 구문패턴에 대해 패턴 번역을 수행하며;
상기 부분 구문패턴의 번역결과를 문장 심볼 S로 치환하고, 그 패턴치환으로 축소된 구문패턴에 대해 패턴 번역을 수행하며;

상기 구문패턴 축소에 의한 패턴 번역이 실패할 경우, 각 구문요소별로 번역을 수행하여 최종 번역결과를 생성하는; 것을 특징으로 하는 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 장치.

【청구항 6】

제 5항에 있어서, 상기 부분패턴 번역부는,

상기 하위절에 대한 부분패턴 번역이 실패할 경우, 다시 절구조 분석 결과를 참조하여 상기 하위절의 하위절에 대한 부분 패턴 번역을 하는 하향식 부분 패턴 번역을 수행하는 것을 특징으로 하는 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 장치.

【청구항 7】

입력 원문에 대해 형태소 분석 및 전처리 청킹을 수행하고, 그 분석결과에 대해 태깅하는 형태소 분석 및 태깅 단계;

상기 태깅결과에 대해 구문 분석을 하고 파싱트리를 출력하는 구문 분석 단계;

상기 파싱트리에서 동사의 하위범주에 속하는 구들의 청킹 결과만을 추출하여 구문 패턴을 생성하는 구문패턴 생성 단계;

번역 패턴을 이용하여 상기 구문 패턴에 대한 번역을 수행하는 구문패턴 번역 단계;

상기 구문 패턴에 대한 번역 패턴 매칭에 실패한 경우, 구문의 절 단위 구조를 분석하는 절구조 분석 단계; 및

상기 절 구조 분석 결과를 참조하여 번역 실패 노드의 하위절에 대한 부분 구문패턴을 생성하고, 그 부분 구문패턴에 대한 패턴 번역을 수행하고 이를 조합하여 최종 번역결과를 출

력하는 부분패턴 번역 단계;로 구성되는 것을 특징으로 하는 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 방법.

【청구항 8】

제 7항에 있어서, 상기 구문 분석 단계는, 입력 문장의 길이가 특정값 이상인 장문일 경우,

분할점 구문단서, 본동사 존재 여부, 분할문의 길이를 바탕으로 2 내지 3개의 분할점 후보를 선정하고;

각 후보별로 그 분할문들에 대한 파싱을 수행하며;

각 분할문 파싱결과에 파싱 가중치를 적용하여 최적의 분할점을 선정하고, 선정된 분할점에 따른 구문 파싱결과를 출력하는; 것을 특징으로 하는 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 방법.

【청구항 9】

제 7항에 있어서, 상기 부분패턴 번역 단계는,

상기 절 구조 분석 결과를 참조하여, 번역실패 노드의 하위절에 대한 부분 구문패턴을 생성하고, 그 부분 구문패턴에 대해 패턴 번역을 수행하며;

상기 부분 구문패턴의 번역결과를 문장 심볼 S로 치환하고, 그 패턴치환으로 축소된 구문패턴에 대해 패턴 번역을 수행하며;

상기 구문패턴 축소에 의한 패턴 번역이 실패할 경우, 각 구문요소별로 번역을 수행하여 최종 번역결과를 생성하는; 것을 특징으로 하는 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 방법.

【청구항 10】

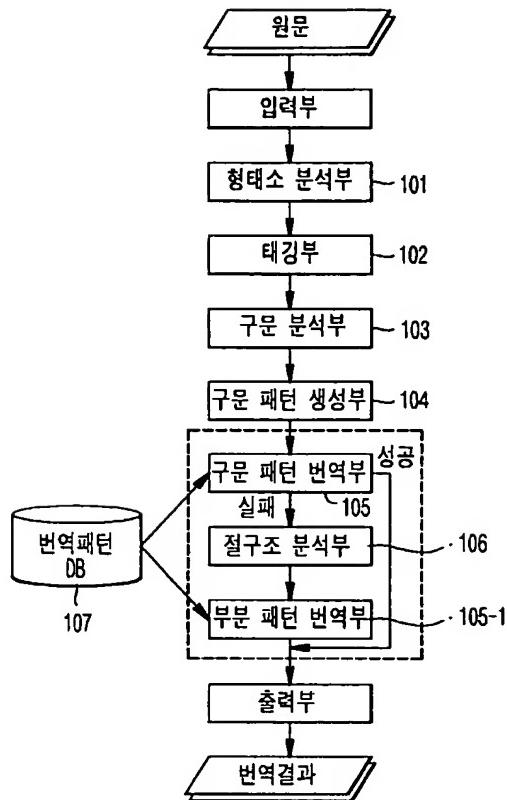
제 9항에 있어서, 상기 부분패턴 번역 단계는, 상기 하위절에 대한 부분패턴 번역이 실 패할 경우, 다시 절구조 분석 결과를 참조하여 상기 하위절의 하위절에 대한 부분 패턴 번역을 하는 하향식 부분 패턴 번역을 수행하는 것을 특징으로 하는 규칙 기반 방식과 번역 패턴 방식을 혼합한 하이브리드 자동 번역 방법.

【청구항 11】

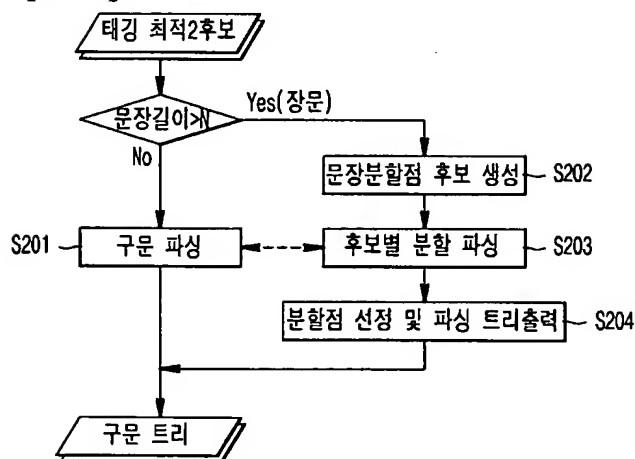
제 7항 내지 제 10항 중 어느 한항에 기재된 규칙기반 방식과 번역 패턴방식을 혼합한 하이브리드 자동 번역방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

【도면】

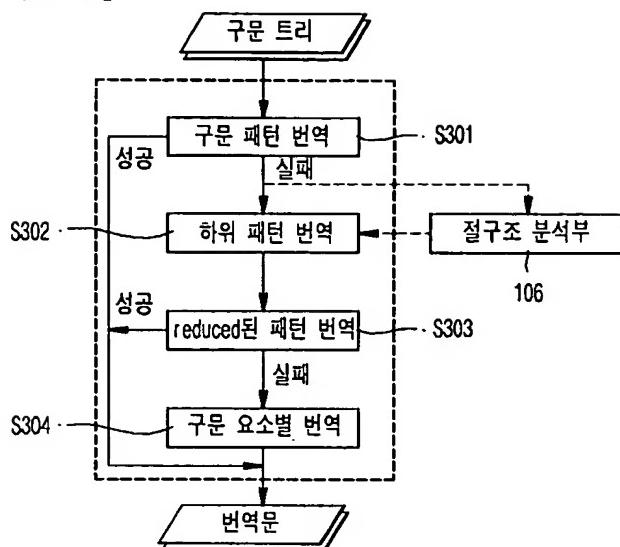
【도 1】



【도 2】

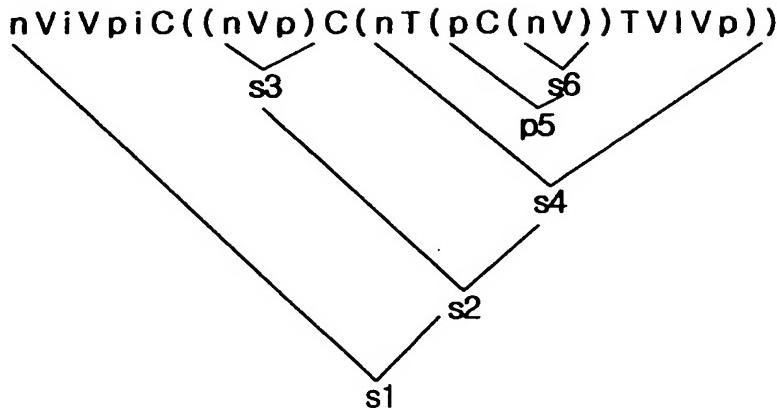


【도 3】



【도 4】

질구조 분석



부분 패턴 번역

1. nViVpiC(nVpCnTpCnTVIVp) → s1

1.1 (nVp)C(nTpCnTVIVp) → s2

1.1.1 nVp → s3 [러시아인이 나토 통제 구조에 일시적으로 참여하]

1.1.2 nT(pCnV)TVIVp → s4

1.1.2.1 pCnV → p5 [우리는 그들이 오늘 말할 때 2 대통령을 포함하여]

1.1.2.2 nTpTVIVp → s4 [(우리는 그들이 오늘 말할 때 2 대통령을 포함하여),

정치적인 지도자가 평화 유지군에서 더 훌씬 광범위하고

러시아 참여를 위해 정돈을 수행하려고 노력하]

1.1.3 sCs → s2 [(우리는 그들이 오늘 말할 때 2 대통령을 포함하여, 정치적인 지도자가

평화 유지군에서 더 훌씬 광범위하고 러시아 참여를 위해 정돈을 수행하려고

노력하는 동안 (러시아인이 나토 통제 구조에 일시적으로 참여하)]

1.2 nViVpiCs → s1 [(우리는 그들이 오늘 말할 때 2 대통령을 포함하여, 정치적인 지도자가

평화 유지군에서 더 훌씬 광범위하고 러시아 참여를 위해 정돈을 수행하려고 노력하는

동안 러시아인이 나토 통제 구조에 일시적으로 참여할) 발표를 기다리라고 듣습니다.]